1/24/86
CP&SD-M

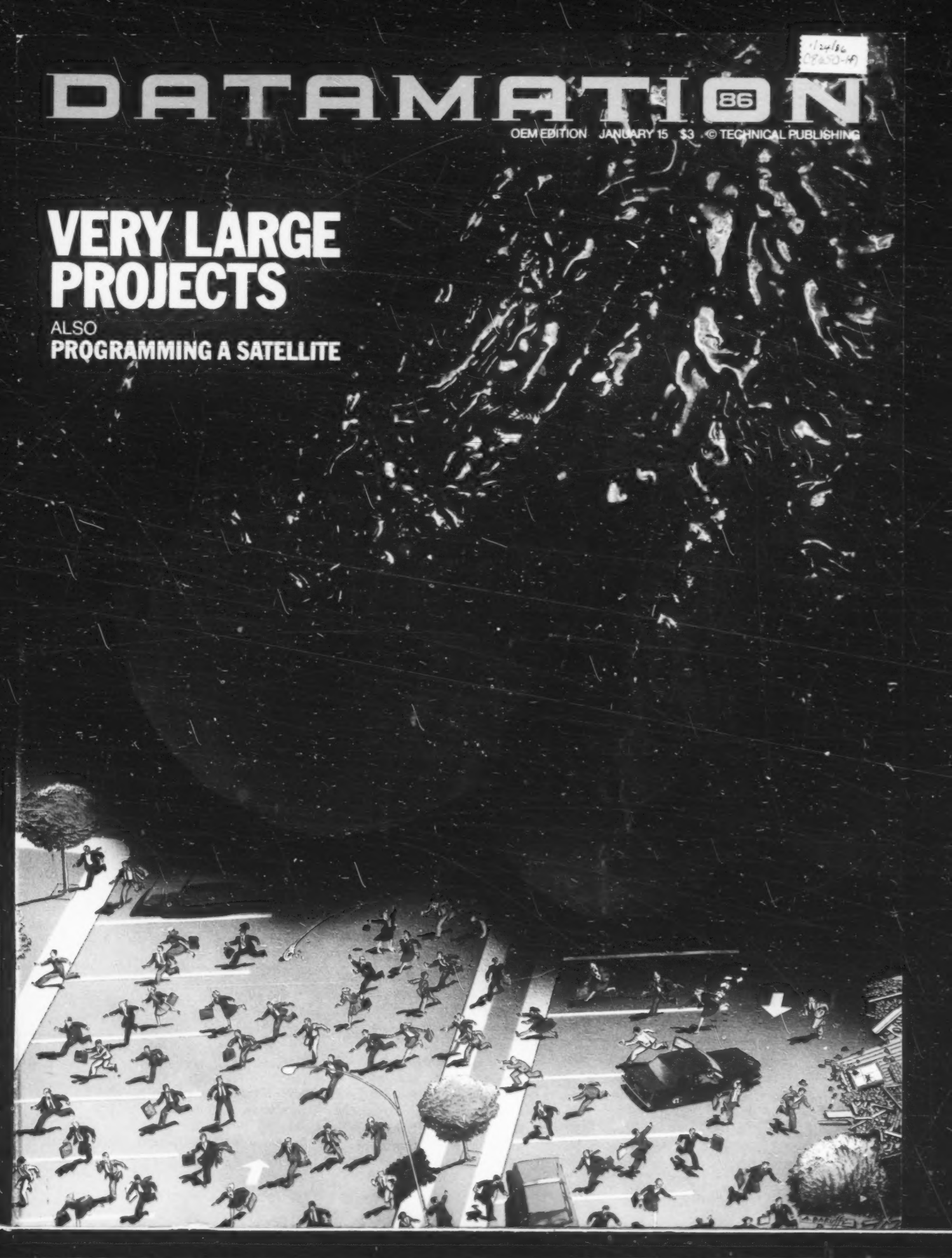
DATAMATION 86

OEM EDITION JANUARY 15 $3 © TECHNICAL PUBLISHING

VERY LARGE PROJECTS

ALSO

PROGRAMMING A SATELLITE



DATAMATION

JANUARY 15, 1986/$3 U.S.A.
VOLUME 32 NUMBER 2
This issue, 191,831 copies

FEATURES

62
VERY LARGE PROJECTS
Daniel S. Appleton
Some systems are too big to be designed using top-down decomposition and implemented using bottom-up structured program modules, and will not move, as a whole, from development into maintenance. So what do you do with them?

75
THE RIGHT STUFF
Joel S. Zimmerman
The successful development of a complicated software system may hinge on learning from previous successes. The development of this NASA satellite software system is a prime example.

85
BOEING'S PC PRACTICES
Richard Metz
Faced with administering five major data centers, more than 5,400 micros, hundreds of minis, in addition to 40,000 software packages, the company has promulgated corporate information systems standards that are intended to bring a focused, integrated approach to its computing activities.

91
NLP TAKES OFF
Tim Johnson
All signs point to the fact that natural language processing is gaining speed as a commercial computing technology.

NEWS IN PERSPECTIVE

26 **TAXATION**
The next step.

28 **RESEARCH**
Doing it at home.

28 **TECHNOLOGIES**
A natural fit?

30 **COMMUNICATIONS**
Airing the issues.

38 **LEGAL AFFAIRS**
When dp needs a lawyer.

45 **COMPUTERS**
Making a dent.

50 **DISK DRIVES**
Testing the waters.

54 **BENCHMARKS**

56 **WORLDWATCH**

DEPARTMENTS

9 **LOOK AHEAD**
14 **CALENDAR**
19 **LETTERS**
23 **HARDCOPY**
95 **HARDWARE**
99 **SOFTWARE**
104 **ADVERTISERS' INDEX**
105 **READERS' FORUM**
107 **SUBJECT INDEX**

OEM SUPPLEMENT 60-1

- 3 **INSURING THE FIT**

COVER PHOTOGRAPH BY TOGASHI. SCULPTURE BY NICK ARISTOVULOS. ILLUSTRATION BY ANDREA BARUFFI

Natural language processing has been more dreamed about than realized, but the optimistic talk is beginning to prove true.

NLP TAKES OFF

by Tim Johnson

After years of looking as though it might never leave the runway, natural language processing is approaching takeoff as a commercial computing technology. The signs are multiplying: the research people are demonstrating new solutions to some old problems; new ventures and new products are bringing these solutions to market; and sales of the products already available are starting to increase. These preliminary successes are stimulating more spending on R&D—and providing the positive feedback that may finally get the whole show off the ground.

What is natural language processing, anyway? Briefly, it is building computer systems to understand English—or any other "natural language"—as it is spoken or written by real people. This means building systems that are able to analyze a string of words to identify its meaningful components and the relationship between them—what the grammarians call parsing.

Parsing is the essential feature of a natural language processing (NLP) system, and systems that do not parse ordinary language are not NLP. This excludes voice recognition systems, which recognize single words or phrases and map them directly onto computer commands. It also excludes "English-like" fourth generation languages: even though individual words may look familiar, they have to obey a formal computer grammar that is not at all that of ordinary language.

Why should computers need to parse natural language? Either because people want to communicate with the machine, or because the machine can help people to communicate with one another. If managers want easy communication with corporate databases on mainframes, NLP can provide a natural language database interface. If a company wants its technical manuals to be read by German (or French, or Spanish, or Japanese) speakers, an NLP machine system can help translate them more quickly.

Many NLP researchers believe their technology will come into its own only with the advent of the "talkwriter"—the machine that converts spoken dictation into written text or database commands. The talkwriter is much closer to realization now than seemed possible a year or two ago.

IBM has already demonstrated a research system with impressive performance. Kurzweil Applied Intelligence of Waltham, Mass., a small startup backed by Xerox and Wang, is proposing to install the first demonstration/evaluation test units of its own talkwriter any month now. The machine is supposed to have a vocabulary of 10,000 words. That would indeed be an impressive and exciting achievement, perhaps even the beginning of a new era in the use of computers. But no one should expect those first test samples to lead quickly to robust operational products. If there is one lesson to be learned from the past history of NLP developments, it is that bringing them up to commercially acceptable levels of performance is a long, slow process.

The two areas where NLP products have gone through this process and achieved commercial acceptance already are mainframe interfaces and machine translation. It took the pioneer of NLP interfaces, Artificial Intelligence Corp. (AIC), also of Waltham, a full six years of development before the first commercial installation of its Intellect system was accepted in 1981. Machine translation systems have been around for even longer, but until very recently they ran only on large mainframes to serve customers with huge translation needs—customers such as the Air Force and the Commission of the European Common Market.

In 1983, IBM broke all precedents by signing an agreement with AIC to sell the Intellect system. Since then, Mathematica Products Group in Princeton, N.J. (part of Martin Marietta Data Systems), has come out with an NLP interface for its Ramis II

CALLIGRAPHY BY RAY BARBER

No one should expect those first test samples to lead quickly to robust operational products.

DBMS. Information Builders Inc., New York, has countered with a version of Intellect that is integrated with its Focus DBMS—Ramis II's biggest competitor. Other companies jumping into the business are Microdata, Irvine, Calif. (part of McDonnell Douglas), and Frey Associates, Amherst, N.H.

SIGNS ARE GOOD

The competition may have done AIC more good than harm. The company shows signs of brightening up its uninspired marketing approach and it reports an increased rate of sales, from a rather steady 70 installations a year to twice that between mid-1984 and mid-1985. Perhaps more important, it can point to a growing number of customers who have learned to use and rely on Intellect, and who can demonstrate some real added value as a result.

Take PPG Industries, for example. Formerly Pittsburgh Plate Glass, this giant with 37,000 employees has its headquarters in a glass-clad Gothic skyscraper in Pittsburgh's Golden Triangle. Since 1983, it has been using Intellect to ease marketing, order processing, and stock control inquiries for the group that handles its fiberglass products. By 1985, Intellect was handling an estimated 300 to 500 inquiries a day, generated both by headquarters staff and by 35 on-the-road salesmen who communicate through portable terminals.

Tim Schell, supervisor of sales and marketing systems, who chose the system and was responsible for getting people to use it—this technology still needs an internal champion—estimates that processing inquiries are between 50 cents and $20 in allocated computer costs. They range from the routine check on the latest sales figures to quite complex impromptu requests. The day before I visited PPG's operation, two staff members had wanted to know what orders had been placed with a certain factory for each week in the month ahead, in order to head off a threatened overstocking problem. "It took two questions and about five minutes through Intellect to come up with data which would have taken hours by any other route," reported one satisfied user.

Machine translation (MT) is following a parallel line of development. Here, while the older established suppliers have stayed with mainframe systems, a new generation of companies has started to offer machine translation systems that run on office minis, or even big micros. Logos, also of Waltham, originally developed its MT skills translating helicopter maintenance manuals during the Vietnam war. It has a German-to-English system that runs on the Wang Office Information System. ALPS (Automated Language Processing Systems) of Provo, Utah, and Weidner of Northbrook, Ill., both grew from the translation efforts of the Mormon Church, and currently offer systems running on a Data General mini and an enhanced IBM PC XT, respectively. Even though they all originated in the U.S., these companies have discovered that the big market for translation is in Europe, especially the in-house translation departments of major electronics and computer companies.

MT SEEMS TO BE A WINNER

There is growing evidence that MT can double the productivity of translators, providing a reasonable first draft of a technical manual for a human translator to correct and polish. MT can also improve quality in some respects, such as consistency of vocabulary. When Veronica Lawson, a leading translation consultant, visited Logos users in the early days of its German-to-English MT system, she found them reporting anything from 20% to 140% gains in productivity. ITT Europe claims "a two- to fourfold increase in output per translator" with its Weidner MT system. ALPS quotes Texas Instruments as saying that after one month of usage, two ALPS terminals had brought down the total costs of handling translations to $25 a page from $55.

As the evidence of success grows, more new products are appearing. There are one or two products offering micro versions of database interfaces; Clout, from Microrim of Bellevue, Wash., is the best known and there are others on the way. But the pc's power in the natural language arena is thus far strictly limited by memory and processing power.

Texas Instruments has a different approach, one that gives the user all the grammatical constructions that the underlying application is able to deal with, but expresses them in preselected words. Using this product, called NaturalLink, the user builds up a query sentence from an "intelligent menu" of words and phrases. The "boilerplate" menu changes dynamically, to show the user all the options available, and it can readily be translated into any language chosen by the system developers. This approach may well prove the best way of providing natural language on micros, at least until they become many times more powerful than they are today.

At the other end of the computing scale, natural language interfaces to mainframes are the subject of major research programs now beginning to bear fruit. SRI International, the nonprofit research center based in Palo Alto, recently completed its Team database interface, funded by DARPA at a cost of over $1.5 million. Compared with existing commercial NLP interfaces, Team will offer wider linguistic coverage, easier portability, greater independence of the underlying database structure, and new facilities for acquiring linguistic knowledge from nonspecialists. Team will be offered on license to universities and research labs, but it will not be turned into a commercial product.

On the other side of the continent, however, BBN, the Boston-based systems house, has announced it will sell a version of its Irus NL interface in 1986. Like SRI, BBN has had major DOD and DARPA contracts for NLP development for many years. Irus shares many of the origins of Team, and will be brought to market with the help of a $5.3 million R&D partnership with Pru-Tech Research and Development of London. Products like Irus should be able to raise the performance of NLP interfaces to a new level; but no one should underestimate the time and effort required to get new natural language products into successful operation in the commercial environment. So far, the six years it took to get Artificial Intelligence's Intellect to market seems the rule rather than the exception.

The other major research objective in this area is the development of dialog interfaces—systems that can support something akin to a two-way conversation with the user, rather than simply responding to individual commands. This type of interface will need to maintain an internal model of the user's objectives and state of mind as the conversation proceeds, something we do unconsciously in human conversation. The artificial intelligence community sees dialog interfaces as an essential for future user-friendly interfaces with expert systems, or for such applications as intelligent computer-aided instruction.

A significant project combining talkwriter-like speech recognition and user modeling to interface to an expert system is Vodis, backed by the government-funded Alvey Program in the U.K. Vodis is intended to provide an automated system for answering telephone inquiries in a limited domain—initially train timetables. It is being developed by British Telecom and a team led by Steve Young, a lecturer in the electrical engineering department at Cambridge University, at a cost of $2.8 million by 1988.

While NL database interfaces and machine translation products have been around for some time, other NLP applications are only now appearing or are still in

FIG. 1

PROJECTED SALES OF NLP APPLICATIONS IN THE U.S. ($ MILLIONS)

APPLICATION	1985	1990	1995
Mainframe interfaces	9	59	184
Micro interfaces	3	120	250
Dialog interfaces	0	8	103
Content scanning	1	45	94
Text editing	0	38	295
Machine translation	2	53	84
Talkwriters	0	95	457
TOTALS	**15**	**418**	**1,467**

Note: Talkwriter projections include dedicated hardware; other figures are for software and support only.

Source: Ovum

the future. The one that looks like the best prospect for an early return has been dubbed content scanning. This means using an NLP system to scan through messages and map them onto a standard format, so they can be loaded to a database or routed to the right person.

Much of the research in this area has been done for the U.S. Navy on hard-copy, ship-to-shore messages relating to equipment faults or enemy contacts. But the most promising commercial application is for processing international money transfer messages for banks.

Several companies are known to be working on content scanning for banks. The first one to announce a product is Cognitive Systems of New Haven. Cognitive's Atrans needs to know a large number of banks by their nicknames; it has to be able to take a telex in French and reformat it in English; it must understand which five-figure number is a security code, which a date, and which a sum of money.

TEXT EDITING PRODUCTS

Products are also expected to appear soon for another application, text editing. This is the use of NLP capability to monitor and correct the style and grammar of written English, as in business letters and other company documents. It could be seen as the natural next step after word processing spelling correctors. IBM's Yorktown Heights, N.Y., laboratory has developed a text editor (they prefer to call it a text critiquer) called Epistle, which is on field test within the company and at a graduate school. The main obstacle to a product launch is probably economics rather than performance; Epistle runs on a mainframe under VM, which would break most people's word processing budgets.

All these applications—interfacing, machine translation, content scanning, text editing—can operate on natural language entered at a keyboard, or already available from a word processor. Clearly, their value would be much increased if they could accept continuous speech as well, which is why so many research teams around the world are racing to produce the first effective talkwriter.

Besides IBM and Kurzweil, already mentioned, another startup, Speech Systems of Tarzana, Calif., has promised to deliver its first sample products before the end of 1986. DARPA is funding research projects at MIT and Carnegie-Mellon University; the British government's Alvey Program is backing a project led by Britain-based Plessey and Edinburgh University; there are at least two projects under way in the big Japanese laboratories at NEC and NTT.

The competing teams will have to combine the very highest skills in signal processing, electronic design, mathematics, acoustics, and human factors, as well as NLP. And it appears that the integrated solution to all these problems is now being rapidly approached.

Yet some forecasts for the short-term sales of talkwriters are entirely too optimistic. At this moment, nobody in the world knows how talkwriters will be used in real offices. It is bound to take several years to solve these problems in ways that will be acceptable to large numbers of users. Once that has been done, the talkwriter will become a very significant product, as indicated in Fig. 1.

NLP SALES SHOULD GROW

The figure shows projections for sales of talkwriters and the other NLP applications discussed here, as prepared for Ovum, a U.K. consultancy specializing in market and technology forecasting in the field of artificial intelligence and advanced digital communications. They must be treated with caution—there is simply not enough information available to produce proper forecasts yet. But they reflect Ovum's view of the relative importance of the different applications, and the possible time scale of their development. They look forward to a time 10 years hence—it is unlikely to be sooner, it could take longer—when the talkwriter will increasingly be seen as the keystone of integrated networks of NLP applications, at least in the larger and more advanced organizations.

Imagine, then, the business environment of the 21st century, when every ambitious multinational has its own corporate language system. Microphones are carefully placed wherever there is an information-gathering requirement: in executive workstations, at conference tables, with the order clerks. Speech is transcribed into text or computer commands. Text editing suggests grammatical corrections and ways to tighten up loose style. Where required, documents are routed on for machine translation, and out across electronic mail networks. Content scanning decides how to index each new document, and how to route incoming messages. Company executives and, where appropriate, outside customers and suppliers, can interrogate the company's databases and other information resources directly. An interlinked hierarchy of lexicons maintains consistency in the corporate use of language, and authorized trade names and new usages are propagated automatically through the system.

An Orwellian vision? Technically impossible? An uneconomic white elephant? Time will tell. To talk of integrating NLP applications when they are barely beginning to prove themselves on a stand-alone basis may seem premature. But the logic is strong. Development of each application will be a lengthy and expensive business. In each case, much of the effort will go into building up vocabularies and defining semantic relationships. Although each individual application will cover a limited domain, there will be extensive overlap between them. Unless they are harmonized and coordinated, chaos may well result.

So, will "language servers" be just one of the standard facilities on the local area networks of the future? Or will the richness and variety of human language—and the thoughts and feelings it tries to express—prove too elusive for computer technology? There will always be a depth and meaning in human communication that will be beyond the power of any computer to follow. But it is a fair bet that where the talk is of stock levels, sales figures, and letters of credit, the computer will more than hold its own.

Never mind the paperless office; in another 30 years, will it be the peopleless office? ◉

Tim Johnson of Ovum Ltd. specializes in the commercial applications of AI technology. He lives in London and has worked as a principal consultant with Logica and for the London *Sunday Times*.